

CLUSTERING K-MEANS PADA DATA EKSPOR (STUDI KASUS: PT. GAIKINDO)

Wakhid Afifi

Fakultas Sains dan Teknologi, Program Studi Sistem Informasi
Universitas Islam Negeri Syarif Hidayatullah Jakarta
Email: wakhid.afifi16@mhs.uinjkt.ac.id

Dhiya'an Ramadhanty Nastiti

Fakultas Sains dan Teknologi, Program Studi Sistem Informasi
Universitas Islam Negeri Syarif Hidayatullah Jakarta
Email: dhiya'an.ramadhanty16@mhs.uinjkt.ac.id

Qurrotul Aini

Fakultas Sains dan Teknologi, Program Studi Sistem Informasi
Universitas Islam Negeri Syarif Hidayatullah Jakarta
Email: qurrotul.aini@uinjkt.ac.id

ABSTRAK

PT Gaikindo adalah sebuah perusahaan otomotif dan produsen kendaraan bermotor yang mengirimkan produknya hingga ke luar negeri. Berdasarkan data ekspor mobil pada PT Gaikindo, terdapat banyak mobil dengan jumlah ekspor yang bervariasi. Oleh karena itu, PT. Gaikindo membutuhkan data mobil mana yang terlaris dan yang tidak laris untuk mengatur volume pengiriman mobil ke luar negeri. Penelitian ini menggunakan metode *clustering* dengan algoritma *k-means*. Metode ini mengelompokkan data dengan cara memasukkan data *brand* dan tipe mobil yang dikirim pada setiap negara. Hasil penelitian ini menunjukkan terdapat 3 *cluster* yang terbentuk dengan kategori sangat diminati, cukup diminati, dan sedikit diminati.

Kata kunci: *clustering; algoritma k-means; data ekspor.*

ABSTRACT

PT Gaikindo is an automotive company and motor vehicle manufacturer that sends its products overseas. Based on car export data at PT Gaikindo, there are many cars with varying export quantities. Therefore, PT. Gaikindo requires data on which cars are best-selling and which are not best-selling to regulate the volume of car shipments overseas. This research uses clustering method with k-means algorithm. This method classifies data by entering brand and type of car data that is sent to each country. the results of this study indicate that there are 3 clusters formed under the category of high demand, quite demand, and little demand.

Keywords: *clustering; k-means algorithm; export data.*

1. PENDAHULUAN

Kegiatan ekspor merupakan sebuah transportasi barang dari negara ke negara lain. Perusahaan dari skala bisnis kecil hingga menengah melakukan kegiatan ini sebagai strategi utama dalam menghadapi persaingan global [1]. PT Gaikindo merupakan sebuah perusahaan agen pemegang merk (APM) yang beranggotakan produsen kendaraan bermotor, distributor kendaraan bermotor, serta perusahaan pembuat komponen utama kendaraan bermotor (*manufacturer*).

Untuk bisnis di dalam negeri, PT Gaikindo memberikan fasilitas kepada para anggotanya dengan kebijakan-kebijakan pemerintah yang menyangkut industri otomotif. Sedangkan di luar negeri, PT Gaikindo aktif berperan sebagai mitra bagi asosiasi-asosiasi industri otomotif yang ada di berbagai negara. Dalam hal ini, PT Gaikindo belum memiliki *insight* mengenai pasar di luar negeri karena pengolahan data ekspor hanya menggunakan *excel*.

Cara untuk mengetahui *insight* mengenai data ekspor tersebut salah satunya adalah dengan mengelompokkan region penjualan menggunakan metode *clustering k-means*. Melalui metode *clustering* dengan algoritma *k-means*, data dapat dikelompokkan berdasarkan pola-pola distribusi, dan dapat menemukan keterkaitan yang menarik antara atribut data [2].

Pada penelitian [3] penerapan algoritma *k-means* dengan inisialisasi *centroid* dengan pengukuran *Jaccard Distance* untuk pengelompokan dokumen-dokumen pada jurnal ilmiah dapat menunjukkan

bahwa sebelum melakukan klasifikasi akurasi meningkat hingga 0,5% dan *recall* meningkat sebesar 0,4% pada artikel yang digunakan. Penelitian ini juga membuktikan bahwa klasifikasi akan mencapai akurasi terbaik dipengaruhi dengan variasi data artikel yang digunakan.

Di bidang kesehatan, penelitian [4] mengenai visualisasi menggunakan algoritma *k-means* pada hasil pemeriksaan laboratorium Klinik Parahita. Hasil penelitian ini menunjukkan bahwa dengan adanya visualisasi dapat memudahkan pasien dalam melihat tinggi rendahnya tingkat nilai pemeriksaan laboratorium.

Penelitian [2] pada sebuah swalayan menunjukkan bahwa proses *cluster* dengan metode *k-means* menghasilkan sebuah informasi mengenai pemetaan penjualan secara terkelompok antara produk yang laris dengan yang tidak laris.

Tujuan dari penelitian ini adalah memberikan gambaran dan solusi kepada PT. Gaikindo dalam menentukan pasar yang terbaik untuk mengeksplor produk mobilnya dengan memaksimalkan penggunaan metode *clustering k-means*.

2. METODOLOGI PENELITIAN

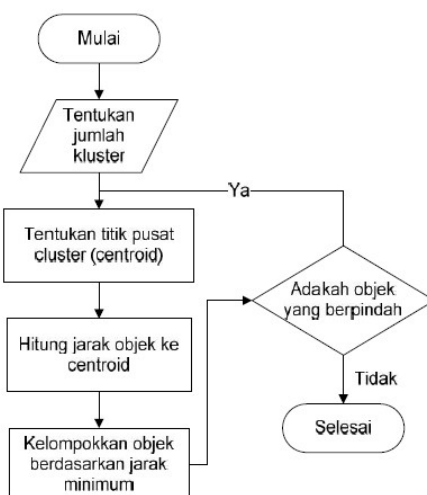
Penggunaan algoritma *k-Means Clustering* dalam penelitian ini bertujuan agar perusahaan dapat menemukan kelompok terbaik. Pengelompokan ini nantinya digunakan untuk melihat potensi negara-negara tujuan ekspor terbaik dengan barang yang paling diminati di negara tersebut, sehingga perusahaan dapat mengoptimalkan proses bisnisnya dengan cara memproduksi apa yang paling diminati dan mengurangi produksi barang yang kurang diminati nantinya.

Proses *clustering* data dengan menggunakan algoritma *k-means* diawali dengan menentukan jumlah *cluster* (*k*) yang diinginkan. Tahap selanjutnya adalah menentukan titik *centroid* C_j . Umumnya, titik *centroid* C_j terbentuk secara bersamaan sesuai dengan jumlah *cluster* yang sebelumnya sudah ditentukan. Setelah *k centroid* sudah terbentuk, langkah selanjutnya adalah menghitung jarak tiap data X_i dengan *centroid* ke-*j* sampai *k*, yang dilambangkan dengan $d(X_i, C_j)$. Untuk pengukuran jarak kemiripan *instance* data menggunakan metode jarak *Euclidean*. Perhitungan jarak *Euclidean* ditunjukkan dengan persamaan 1.

$$d(X_i - C_j) = \sqrt{\sum_{i=1}^N (X_i - C_j)^2} \quad (1)$$

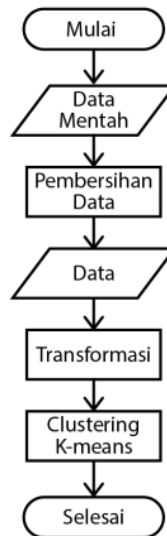
Menurut [5], apabila jarak $d(X_i, C_j)$ semakin kecil, maka kesamaan antara dua unit pengamatan semakin dekat, dengan ketentuan bahwa semua fitur dalam dataset tidak saling berkorelasi. Apabila dataset yang digunakan memiliki fitur berkorelasi, maka perhitungan jarak dapat menggunakan metode *Mahalanobis*.

Tahap selanjutnya menurut [6], adalah mengelompokkan data berdasarkan jarak minimum atau jarak *centroid* yang paling dekat. Tahap berikutnya adalah menghitung rata-rata jarak seluruh data terhadap *centroid*, kemudian titik *centroid* diperbaharui sesuai perhitungan. Setelah itu, akan diulang lagi dari proses awal. Iterasi ini akan terus diulang hingga titik *centroid* sudah tidak berubah lagi. Penjelasan ini ditunjukkan pada Gambar 1.



Gambar 1. Algoritma *K-Means Clustering* [7]

Pengolahan data melalui proses *clustering* dimulai dengan didaptkannya data mentah berbentuk *file excel* yang kemudian dibersihkan agar menghilangkan data yang tidak konsisten. Hasilnya didapatkan data yang bersih. Data ini selanjutnya mengalami proses *transformasi* untuk mengubah data yang tadinya *alphabet* menjadi data angka atau numerik. Tahap terakhir yaitu *Clustering K-means* dengan membagi *cluster* menjadi 3 berdasarkan kriteria produk yaitu sangat diminati, cukup diminati dan sedikit diminati. Alur pengolahan data diilustrasikan pada Gambar 2.

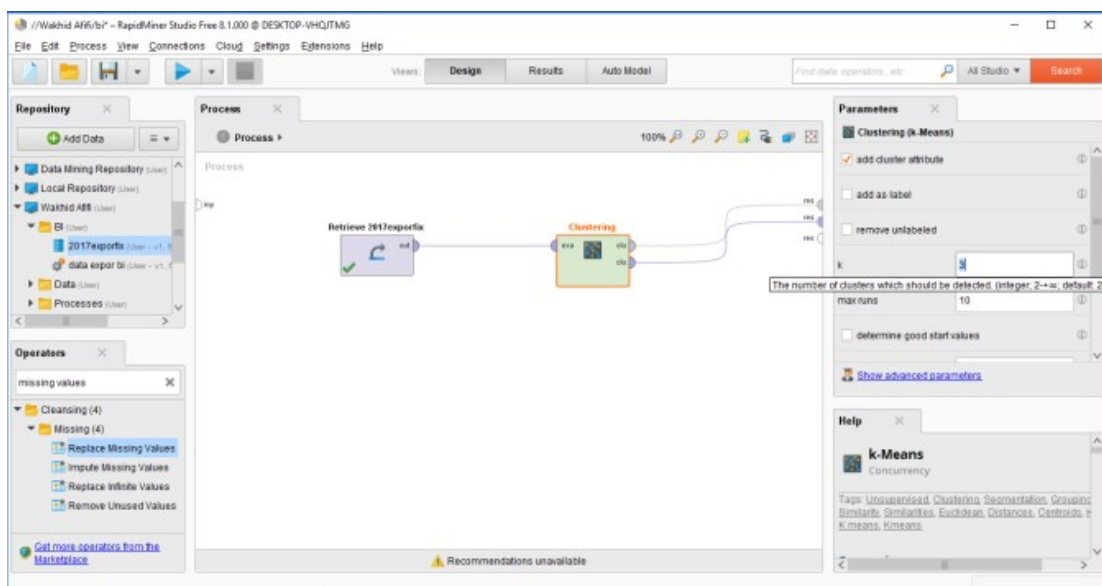


Gambar 2. Alur Pengolahan Data

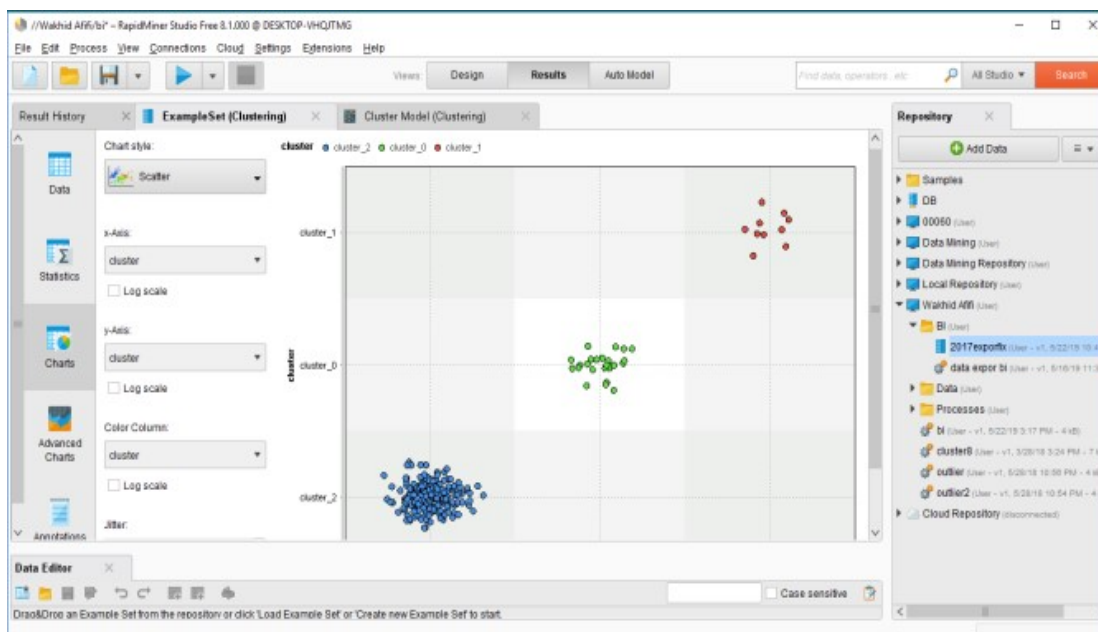
Penelitian ini menggunakan data ekspor PT Gaikindo yang didapatkan dari situs www.gaikindo.or.id. Data yang digunakan ialah data pada tahun 2017 yang sudah mengalami proses pembersihan data.

3. HASIL DAN PEMBAHASAN

Desain *RapidMiner* yang telah dibuat kemudian di-*run* agar dapat diproses untuk menampilkan hasil. Hasilnya didapatkan 3 *cluster* yang memiliki selisih jumlah data yang jauh pada setiap *cluster*. Model ditunjukkan oleh Gambar 3. Hasil ditunjukkan pada Gambar 4 dan 5.



Gambar 3. Model Proses *Clustering*



Gambar 4. Hasil Clustering K-Means

Attribute	cluster_0	cluster_1	cluster_2
brand	1.100	1.700	1.374
Model	2.960	6.900	4.324
country	10.800	15.200	11.927
total 2017	2803.120	7794.400	176.799

Gambar 5. Hasil Clustering K-Means

Centroid yang didapat pada Gambar 4 dapat diartikan:

- Cluster 2: Sedikit diminati (rata-rata pemesanan 176 unit)
- Cluster 0: Cukup diminati (rata-rata pemesanan 2.803 unit)
- Cluster 1: Sangat diminati (rata-rata pemesanan 7.794 unit)

Jumlah masing-masing cluster dari total 254 items data ialah:

- Cluster 2: 219 items
- Cluster 0: 25 items
- Cluster 1: 10 items

3.1 Hasil penelitian dibahas sebagai berikut.

3.1.1 Untuk Brand Toyota

Model Fortuner memiliki sedikit peminat di banyak negara yaitu Bahrain, Jordan, Kuwait, Lebanon, Oman, Ecuador, Brunei, Thailand, Srilanka, Papua, Kamboja, Laos, Myanmar, Qatar, Peru, Singapore, Fiji, Iraq, dan Tahiti. Ini diketahui karena terdapat pada *cluster 2*.

Fortuner memiliki cukup banyak peminat di negara Saudi Arabia karena terdapat pada *cluster 0* dan memiliki banyak peminat pada negara Vietnam dan Filipina karena terdapat pada *cluster 1*.

Model Innova memiliki sedikit peminat di negara Brunei, Fiji, Oman, Argentina, Ecuador, Qatar, Lebanon, Bahrain dan Kuwait karena terdapat pada *cluster 2*. Memiliki cukup banyak peminat pada negara Saudi Arabia dan Thailand karena terdapat pada *cluster 0*.

Model Vois memiliki sedikit peminat pada negara Bahrain, Kuwait, Brunei, Qatar, Singapore dan Iraq karena terdapat pada *cluster 2*. Memiliki cukup banyak peminat pada negara Oman karena terdapat pada *cluster 0* dan memiliki banyak peminat pada negara Saudi Arabia karena terdapat pada *cluster 1*.

Model Yaris memiliki sedikit peminat pada negara Brunei karena terdapat pada *cluster 2*. Model Sienta memiliki sedikit peminat pada negara Brunei, Malaysia, dan Singapore karena terdapat pada *cluster 2* dan memiliki cukup banyak peminat pada negara Malaysia dan Thailand karena terdapat pada *cluster 0*.

3.1.2 Untuk Brand Daihatsu

Model TOWN ACE memiliki banyak peminat pada negara Jepang karena terdapat pada *cluster 1*. dan Model LITE ACE memiliki cukup banyak peminat pada negara Jepang karena terdapat pada *cluster 0*.

Model Gran Max memiliki sedikit peminat pada negara Brunei dan Malaysia karena terdapat pada *cluster 2*.

Model Avanza memiliki sedikit peminat pada negara Bahrain, Kuwait, Lebanon, Oman, Thailand, Srilanka, Kamboja, Laos, Malaysia, Qatar, Iraq, Guatemala, Honduras, Bhutan dan Tahiti karena terdapat pada *cluster 2*.

Memiliki cukup banyak peminat pada negara Peru dan Saudi Arabia karena terdapat pada *cluster 0*, dan memiliki banyak peminat pada negara Filipina dan Mexico karena terdapat pada *cluster 1*.

Model Rush memiliki sedikit peminat pada negara Papua karena terdapat pada *cluster 2*. Model Wigo memiliki sedikit peminat pada negara Brunei, Srilanka, Guatemala, Honduras, Panama, dan Peru karena terdapat pada *cluster 2*. Memiliki banyak peminat pada negara Filipina karena terdapat pada *cluster 1*.

3.1.3 Untuk brand Hino

Model DUT RO memiliki sedikit peminat pada negara Vietnam, Filipina, Kamboja, Laos, dan Myanmar karena terdapat pada *cluster 2*. Dan model XZU memiliki sedikit peminat pada negara Vietnam dan Filipina karena terdapat pada *cluster 2*.

4. KESIMPULAN

Dari permasalahan pengelompokan negara untuk tempat ekspor mobil dapat diselesaikan dengan menggunakan metode *clustering k-means*. Walaupun mampu menyelesaikan data dengan jumlah banyak, namun masih belum efisien dalam mengelompokkan dokumen secara tepat.

Hasil akhir juga sangat bergantung pada *centroid* yang ditentukan, karena jika berbeda *centroid* yang ditentukan maka hasilnya pun akan berbeda juga.

UCAPAN TERIMA KASIH

Terima kasih untuk PT. Gaikindo yang telah memberikan kesempatan kepada kami untuk mengolah data eskpornya. Terima kasih juga kepada semua pihak di Universitas Islam Syarif Hidayatullah Jakarta atas dukungannya terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] H. Deresky, *International Management*, 4th ed. United States Of America: Addison-Wesley, 2006.
- [2] B. M. Metisen and H. L. Sari, "Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila," vol. 11, no. 2, pp. 110–118, 2015.
- [3] A. Yusuf and T. Priambadha, "Support Vector Machines yang Didukung K-Means Clustering Dalam Klasifikasi Dokumen," *JUTI J. Ilm. Teknol. Inf.*, vol. 11, pp. 13–16, 2013.
- [4] Y. Ardy and C. H. Budi, "Visualisasi Hasil Pemeriksaan Laboratorium Pasien Studi Kasus : Parahita Diagnostic Center," vol. 01, no. 02, pp. 142–150, 2015.
- [5] P. L. O. S. Duran, "The Cluster Problem and Preliminary Ideas," in *Cluster Analysis*, Springer, Berlin, Heidelberg, 1974, pp. 1–2.
- [6] Y. Agusta, "K-Means – Penerapan, Permasalahan dan Metode Terkait," *J. Sist. dan Inform.*, vol. 3,

no. Pebruari, pp. 47–60, 2007.

- [7] N. Wakhidah, “Clustering Menggunakan K-Means Algorithm (K-Means Algorithm Clustering).”